Machine Learning – Behind the Hype, and What You Really Need to Know

A Bentley White Paper

Richard Irwin
Senior Product Marketer,
Asset Performance

Published:
October 2018

Bentley®
Advancing Infrastructure

The terms “artificial intelligence” and “machine learning” have been around for decades and mainly crop up in the realms of science fiction, with robots doing the chores around the house or androids taking over the world. The hype and buzz around artificial intelligence and machine learning have never really gone away, and these concepts are now on the verge of reality. Both artificial intelligence (AI) and machine learning (ML) can be applied practically in the modern world – simply look at the advancement of the self-driving car in the past five to 10 years. Business organizations are now actively looking at how machine learning and artificial intelligence can be applied to their businesses to achieve goals and solve challenges previously thought to be unobtainable. However, without the proper knowledge and understanding, machine learning can be full of misleading promises; if an owner-operator does not do the necessary due diligence to learn what machine learning can do and what its limitations are, there may be many frustrations, disappointments, and project failures in the future.

This paper will dispel some of the myths that machine learning has generated over the years to build up expectations and look at some of the most important questions any business looking to implement machine learning should ask themselves to ensure they are prepared and ready for the machine learning revolution.

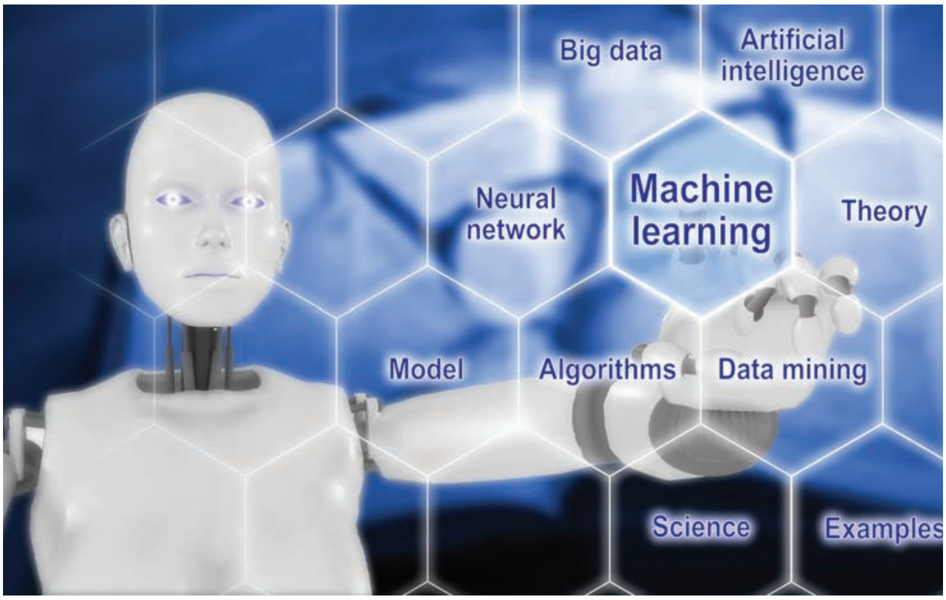
Artificial intelligence is a term that seems to attract most of the unwanted attention when speaking within this space, but it actually is the all-encompassing concept, with everything else below it being a subset of AI.

Myths of Machine Learning

Machine learning, advanced analytics, predictive analytics, artificial intelligence: They are all the same, aren’t they?

With so much excitement surrounding machine learning on the internet and in the meeting room, it is often confusing to determine what people are talking about when the latest buzzwords are used to explain the most recent trends. Confusion occurs when terms are used in error or in the wrong context, especially when many varying terms are used interchangeably to describe the same thing.

This applies to machine learning, predictive analytics, artificial intelligence, advanced analytics, data science, deep learning, and many more. All can be said to enable organizations to make more accurate decisions. While some of these concepts often overlap, they differ from each other in terms of application, what they do, and how they do it. Artificial intelligence is a term that seems to attract most of the unwanted attention when speaking within this space, but it actually is the all-encompassing concept, with everything else below it being a subset of AI.



The idea that you can "throw data at machine learning" or "throw machine learning at your data" and results will be delivered instantaneously is wrong.

To summarize:

- **Artificial Intelligence** – the broader concept of machines being able to carry out tasks in a smart way using cognitive processes via computer programs, from robots to self-driving cars.
- **Machine Learning** – a current technique that enables computers to get into a mode of self-learning with data without being explicitly programmed; the computer uses algorithms to parse data, capture knowledge, learn, and make a predictive model so owners can form a business decision.
- **Predictive Analytics** – uses machine learning algorithms to make predictions about future trends, behavior, and activity. Based on current and historical scenarios, predictive analytics help businesses analyze data to find patterns, which are used to make predictions about future events.
- **Deep Learning** – is a subset of machine learning concerned with algorithms inspired by the structure and function of the brain called artificial neural networks that mimic the human brain's processes.
- **Data Science** – is the collective term for the use of methods, processes, statistics, algorithms, and systems for extracting knowledge, information, and insights from data, usually by data scientists.

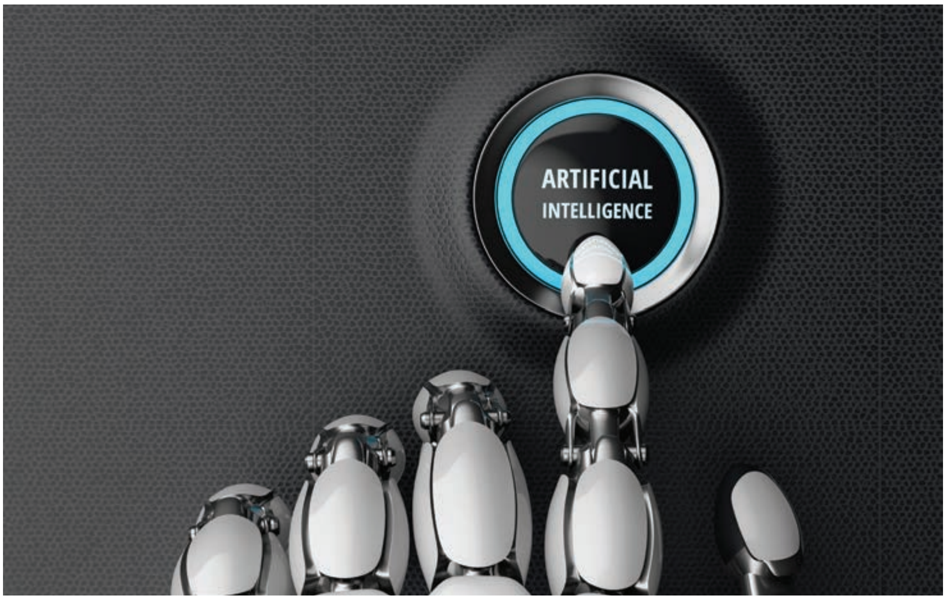
It's important to understand what each subset does and how it can apply to your organization, your data, your business challenges, and, ultimately, your goal. Not every application will suit your challenge, so it's important to investigate further and see which methodology will help you achieve your goals; it could be more than one.

Just push the 'Magic Button' and machine learning will solve everything, right?

There seems to be an impression that with machine learning, or any other form of predictive analytics, all you need to do is press a button and machine learning will automatically be applied to the data and transform it into something magical and insightful. This process would all be carried out within seconds, providing you with the answers and predictions you require. However, it is not that simple.

While the level of training may not be as intensive for standard machine learning projects, there is still a lot of human effort required throughout the process, and that's before we start talking about full autonomy.

Machine learning and artificial intelligence are currently being presented as processes that can solve any problem or situation. Yet, like every new technology, both techniques have limitations, but they can be very useful in certain ways if used correctly. For instance, machine learning and artificial intelligence can provide the ability to self-learn and solve business problems that you were aware of, but unsure of how to rectify. Machine learning is a complex process that requires many elements to work properly to deliver insights. It is not a "plug in and play" application. Time, effort, and a lot of periods of trial and error are necessary to understand the current situation and how to optimize the applications. The idea that you can "throw data at machine learning" or "throw machine learning at your data" and results will be delivered instantaneously is wrong.



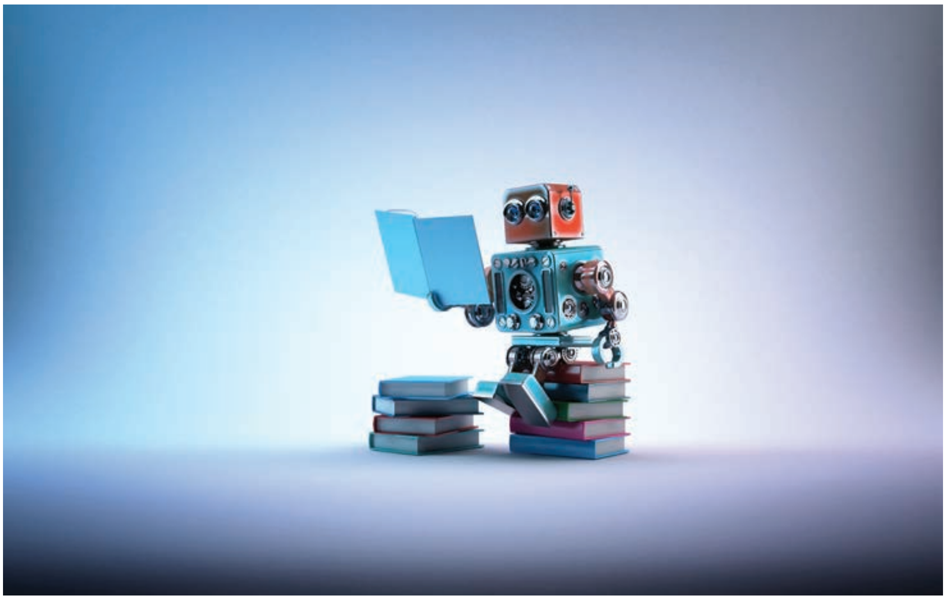
To start, you must have quality data in place. Then, you will select the technique, or techniques, that is appropriate to your business problem. You need the right algorithms to build a model; therefore, you will likely want a data scientist to be a member of your team, working with subject matter experts who understand the business problem. Additionally, you will need supplementary methods and applications and to learn new scripting languages. This all takes time.

Machine learning is still in the very early stages of development, and while clear and accurate results at this stage appear to be limited, those who are educated about the technology understand the effort required to obtain real and usable results. As machine learning matures over time, it is anticipated to revolutionize industry. As our understanding of machine learning develops, we will wonder how we lived without it from a business, industrial, and personal perspective.

Machines can learn autonomously

It is frequently thought that once machine learning has been applied, it will learn autonomously, it will self-learn by itself without any more human intervention. This is wrong on two accounts. First, humans taught the system to self-learn in the first place. Human programmers are still required to orchestrate and design the machine learning's architecture; human engineers and data scientists are still required to feed it the necessary structured data for it to train and recognize patterns and potential events (and if using unstructured data, even more time, effort, and intensive training is required). Secondly, once machine learning is up and running, it still requires humans to maintain it, to create new and update existing algorithms, and to make the decisions based on the results generated. The human influence and effort needed to help with the development of self-driving cars is a concrete example of this. If it were that easy to do, then we would not continue to hear about crashes and driving incidents in the media. The amount of effort required to simulate a human's reactions when driving and the recognition systems that are involved in tagging and recognizing a multitude of different objects and the situations they present on the road are immense. Then, there are ethical decisions to program for the "what if" scenarios. While the level of training may not be as intensive for standard machine learning projects, there is still a lot of human effort required throughout the process, and that's before we start talking about full autonomy.

You first need to ask yourself if you need machine learning or if advancing your analytics or enterprise applications would be enough to solve your business problems.



Here's a month's worth of data, predict the next 10 years!

Data is everything. Data is the most important asset any business owns when it comes to AI and machine learning. Data can change the world, from curing diseases to programming self-driving cars. Across numerous industries, data is transformational and key to notifying engineers about how their assets are performing and when they might fail, providing information on how the overall business is performing financially and operationally at any specific time. Data is the lifeblood of any digital business, stemming from assets, equipment, people, sensors, geographic information systems (GIS), and information technology (IT) applications, to name a few. Without data, any business would not function normally and remain competitive.

The concept of machine learning is becoming a reality due to factors like processing power, analytics, and technology finally reaching the levels of maturity to advance, and data is the most important component. As more vendors adopt machine learning, such as that offered by Bentley Systems' asset performance management solution, ensuring a user has suitable data is of paramount importance. Irrelevant, missing, or unreliable data is useless, especially when used at the start of a machine learning process. Bad data will result in unusable information or inaccurate predictions, so up-to-date and reliable data is key. The amount of data is the most important – the more data provided (specifically the most relevant data), the more accurate the prediction. Potential users of machine learning who have not kept comprehensive data records expect miracles to occur when they arrive at the data scientist's door with a sample, which in reality is the sum total of their operations data. To create accurate insights and predictions, historical data must go back months and years to present a credible picture of performance. In addition, the data must also include a variety of examples or events for the model to learn what previously happened and to accurately predict what could potentially happen in the future.

Preferably, data needs to be prepared and clean before it goes anywhere near a machine learning application. Data characteristics to avoid include:

- Incomplete: Data lacks attributes or contains missing values.
- Noisy: Data contains erroneous records or outliers.
- Inconsistent: Data contains conflicting records or discrepancies.

Despite using digital technologies to observe activities in an industrial setting, giving instructions for changing processes and actions and making sure instructions are carried out with integrity increases the value of machines and decision makers working together.

While there is debate on whether to include all available data sources before adding everything to the mix, it's important to identify what data isn't relevant to the goal and remove it. Once the data is selected, it must be determined if it needs to be processed into the right shape and format, live in a database, or if it would be better in a text file. Additionally, if you have large datasets comprising raw values, they may be easier to use (and reduce the running time of models) if they are aggregated.

These factors, and many more, need to be addressed before any form of algorithm can be created to produce results. Failing to address these issues will only waste resources and time because you only get back what you put in.

Everyone is talking about machine learning, so I must need it!

As mentioned previously, machine learning is generating so much hype that it appears to be the new "must have" technology. The benefits in early adoption case studies do indeed seem to make it worth the effort, but do you really need it? While it promises a lot, do not necessarily assume it will sort out your business problems automatically. You first need to ask yourself if you need machine learning or if advancing your analytics or enterprise applications would be enough to solve your business problems.

It's worth noting that machine learning is still very much in it's infancy, and the results you expect will not materialize the way you want them, in the time you want them. Expectations for machine learning should be relatively low due to its immaturity, especially from an industrial perspective. The accuracy of predictions will take time to reach an acceptable level before they can positively affect a business. For instance, machine learning will not provide you with an accurate estimate that a piece of equipment will fail on this day, on this hour, on this minute.

You can get a reasonable indication that something is trending toward failure using condition-based maintenance and a dashboard showing the trend and alarms with calculations and rules engines built in, as Bentley's AssetWise does. But where there are costly recurring problems on complex systems or components, machine learning is worth investigating. There will be instances of false positives and negatives that could obscure results depending on the level of confidence and probability of the prediction. Therefore, it's best to start small and extend the analytical experience gradually, like using a hybrid version of machine learning with an existing analytical application. Due to machine learning's infancy, the best way to start adopting machine learning is to learn, educate, and understand its value and potential and prepare for it. This means training reliability engineers with new scripting languages, like R or Python, and making sure that the data you have is cleaned up so important features can be easily identified. The next generation of reliability engineers will have many more digital and programming skills.

While the machine learning buzz is all around us, there is no need to fear being left behind. Preparation is key toward useful and meaningful machine learning adoption. It is important to start small with a pilot to learn the techniques and processes involved, reiterate, and learn from mistakes, and then expand onto the next level to achieve real business value.

Machine learning and artificial intelligence will take away our jobs!

There are hotly debated discussions concerning the outcome that machine learning and automation will take over human actions and replace jobs as artificial intelligence advances. There are even calculators available to determine how "at risk" your job is from computers. Throughout history, technological advancement has always caused disruption and fear. During the first industrial revolution, skilled workers, such as weavers, were made somewhat obsolete when automated looms took over their roles, but then other different jobs were required, like managing the machines and ensuring the output met high-quality standards. There are numerous examples over the years of technology taking over industries, such as typing pools being replaced by automated dictation through speech analysis; the invention of the steam engine replacing field workers; and the automated assembly line. Manufacturing has been most affected by automation – simply look at the car industry. Replacing monotonous and repetitious work, robots can perform faster and with more precision than a human, so it is an obvious choice for automation. Repetitive work on an assembly line is better automated since the machine will do the same movement repeatedly once it has been taught and programmed how to do it. And it will continue to do that action all day without tiring or complaining.

> Once data is selected, it could come in many different shapes and forms from multiple systems. It needs to be all in the same format, such as XML, csv, text file, and many more.

Jobs that are repetitive and can be a threat from machines can be divided up into two distinct work groups: cognitive versus manual, or jobs that require thinking versus jobs that do not, respectively. Routine administrative work, with simple processing and calculations, is easier to automate than handling complex customer questions or problems. Big data number crunching, with its volume, variety, and velocity, is the driver behind machine learning, amplified by the Industrial Internet of Things. While it can be said that machine learning in data number crunching is replacing traditional analyst roles because of the speed machines can compute and predict scenarios, it is also creating new roles.

With all the data being generated, there will be more jobs available for data scientists and statisticians as well as subject matter experts in their field to interpret and act on the outputs that the digital data and machine learning will provide. Machines will do the thinking processes, but a human will still need to make the best decisions based on the results. Therefore, decision makers are still an integral part of the chain in digitalization. Despite using digital technologies to observe activities in an industrial setting, giving instructions for changing processes and actions and making sure instructions are carried out with integrity increases the value of machines and decision makers working together.

If your job requires negotiation or a high degree of creativity, there's also less risk of losing your job to a machine. A surgeon requires years of education, training, knowledge, and experience to perform their job accurately. Machine learning can certainly help with diagnosing a medical problem through the analysis of a patients' symptoms, conditions, and medical history, but performing a complete surgery is still far away from what artificial intelligence can accomplish. Robots are currently assisting surgeons who perform invasive procedures, but the robots are controlled by the surgeon.

There are roles that machines and technology cannot replace. Until the time comes when computer intelligence is indistinguishable from human intelligence, it seems the best answer is to embrace the technology and work with it, not against it. Working side-by-side with technology will bring tremendous advantages to businesses, and while some roles will be diminished, other roles will grow or be created. The question will be about how we choose to adapt to technology, automation, and artificial intelligence and make the best out of it.

5 Questions You Need to Ask Before Investing in Machine Learning

Question 1: What do you want to accomplish with ML?

While it is not necessarily a prerequisite to have a data scientist as part of the team, most successful organizations have employed a data scientist. As it stands, data scientists are becoming sought-after assets.

To consider implementing machine learning in your business, you need to ask yourself, why? What business goal are you trying to achieve; what challenge is proving impossible to complete; what is my data not able to tell me; what will machine learning achieve that I cannot achieve currently? You need to start with a question to be sure that there is a problem to be solved. The question will clarify the data you need to collect to form the answer.

From an industrial perspective, the most common problem is typically unsolved recurring equipment failure, the need to predict failures or operational events, and anomaly detection, with the result being an improved service or production cost and quality. Do not implement machine learning to solve all problems because it is generating the most hype within the industry now; make sure you have a genuine issue and a lot of relevant data (likely underused currently) with which machine learning can help you gain quicker responses to early warnings and other key asset performance indicators.

Question 2: Is your data prepared for ML?

Machine learning needs data. Without it, all the work, effort, and costs will be meaningless. Machine learning algorithms model problems with data, and the quality of the modeling process depends on the quality of your data. If your data is incomplete, full of gaps, unreliable, untrustworthy, or not compatible, then it will greatly affect the predicted models. While more data is important, only data relevant to the question or problem should be included. Throwing everything within the database will not help anyone, so be selective. Larger data sets will also take longer to run.

While having a lot of reliable, accurate data is ideal for modeling, if you are trying to predict, for instance, the likelihood of an asset failing, then you also need to have a lot of available data from when things went wrong. To train the model to spot anomalies, the model needs to know what "good" looks like just as much as what "bad" looks like. There may be some instances where little data information is available, such as when reliability failures occur, which are rare, or even when many failures occur, data simply may not be available. In this case, data will need to be created as close to the actual situation to recreate the problem. You may not be able to recreate every eventuality, but the more "right" data you feed the algorithm, the more accurate the model will be.

By preparing yourself at the start of the journey by answering the five questions within this white paper, you will find yourself well on the way to a successful and smoother experience with machine learning.

An important aspect regarding data selection relates to data availability; make sure you have access to the data you require. It could be locked away or contain private information, like financial data, but if it's pertinent to the problem, then insist on its inclusion.

One last element is format and the preprocessing of data. Once data is selected, it could come in many different shapes and forms from multiple systems. It needs to be all in the same format, such as XML, csv, text file, and many more. Ensure the data is clean by removing invalid or erroneous information, fill in any gaps if possible, and make sure they all include timestamps. Once finalized, feature engineering can commence, which includes aggregation, such as taking averages or counts, as opposed to each individual raw value.

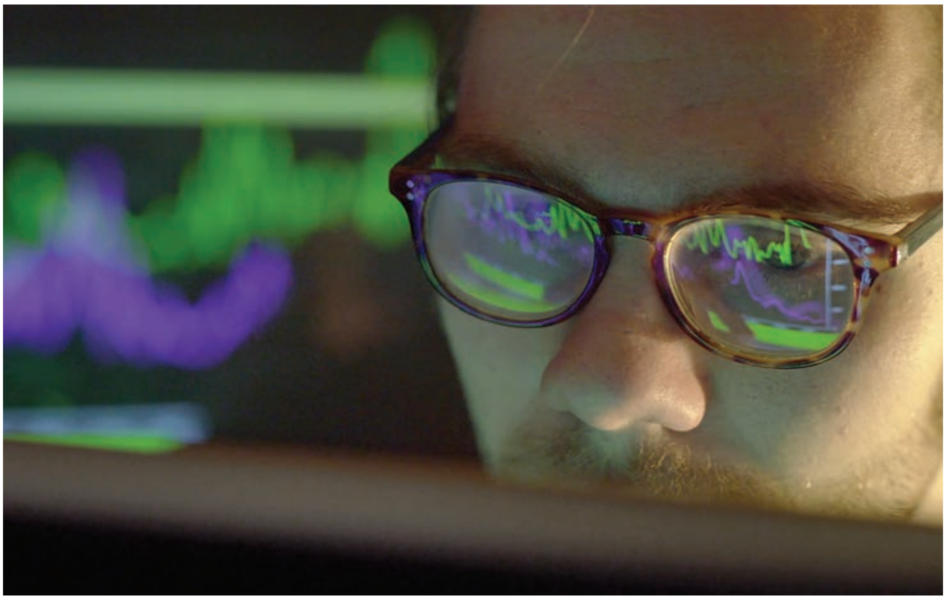
Question 3: Build or Buy?

Once you have discovered a use or opportunity for machine learning in your business, the next question you need to answer is whether you should buy from a vendor or build your own. While it is possible to build your own machine learning platform and it gives you the control to build a platform specific to your needs, this method takes a significant amount of time, specific skills, and investment. Investing in a platform such as Microsoft Azure gives you a trusted and capable solution for a secure, private cloud platform aimed at developers and data scientists who can get started quickly. Purchasing from a vendor is not as simple as it sounds, as you are not buying a "plug in and play" product, and it still requires investments in learning and skills. Similarly, buying or building your own algorithms is a decision that needs to be thought through by considering aspects of time and accuracy.

Building your own machine learning platform may be the best option if you are confident in the skills you have in-house among your designers, engineers, architects, developers, and scientists, and you think you can provide something more to your customers and business. However, like all application building, experience matters when creating an application that exceeds the reliability, stability, knowledge, and integration of other vendor applications.

Question 4: Do you need a data scientist?

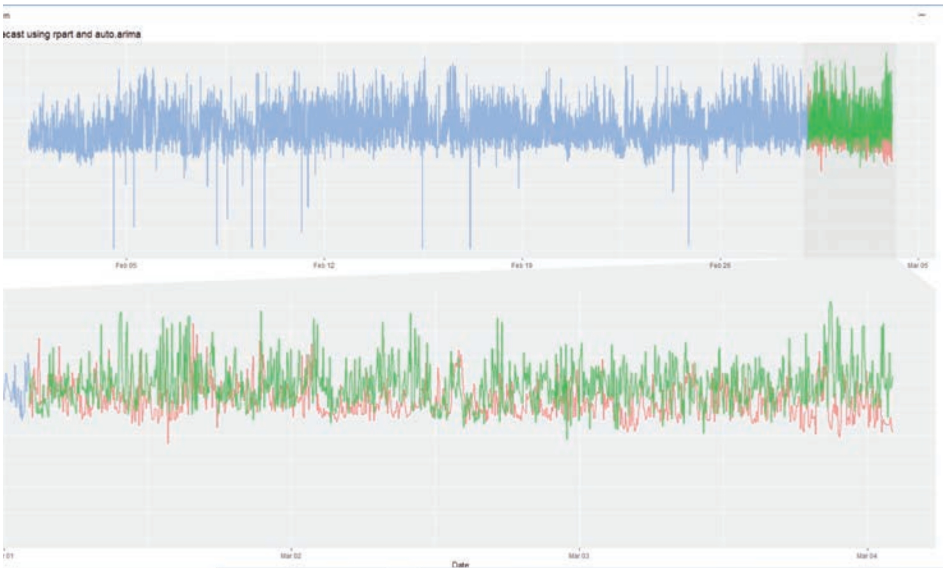
A data scientist can be defined as someone who is employed to analyze and interpret complex digital data to assist decision making, and someone who possesses a multitude of skills and knowledge of mathematics and programming. The terms "data scientist" and "machine learning" are often linked together. Gartner, Inc., a global research company, noted that venturing into machine learning is difficult without data scientists.[1]



While it is not necessarily a prerequisite to have a data scientist as part of the team, most successful organizations have employed a data scientist. As it stands, data scientists are becoming sought-after assets. While in-house engineers and analysts may have knowledge of machine learning, and can buy algorithms off the shelf, having a data scientist on premise will prove less challenging and daunting. A good data scientist will work well in a team by sharing knowledge with in-house subject matter experts, project staff, and IT operations. There will most likely be a need for a data scientist or analyst (data interpretation) with machine learning, but they should be integrated throughout the whole team and work closely with subject matter experts to define and make machine learning part of your new system of intelligence.

Question 5: How do you get the most out of ML?

If you have chosen to acquire a platform for machine learning, the challenge does not stop there. There are different frameworks, libraries, applications, toolkits, and data sets in the machine learning world that can be very confusing, but once understood, they allow anyone to start a journey by a process of trial and error, using machine learning to solve real-world problems. These platforms vary in complexity, user-friendliness, and depth, with each one having their own strengths and weaknesses. Some are hands-free, like Amazon and Azure machine learning, which allow you to dive in straight away without coding in R, Python, and other languages. That is not to say that you cannot still incorporate additional languages or technologies into your machine learning toolkit. It all depends on what is the best fit for your challenges. In addition to different languages, there are different forms of models and algorithms that can be implemented, like neural networks, cluster analysis, decision tree, and more. These relate to the type of machine learning used (supervised or unsupervised) and the type of business objective. Again, research, knowledge, and experimentation need to be applied to ascertain the best methodology for your needs.



AssetWise forecasting example showing data modeling technique, displaying the accuracy of the forecasted model against the actual data.

Research into multiple techniques need to be undertaken to assess your needs against their offerings. Some will use different approaches and techniques, such as retail versus industrial, big data versus industrial analytics, or single variable or multi-variant. Some platforms will have different data requirements, like only working with clean or raw data.

Conclusion

Once you have asked these questions and answered them, then comes the hard work of implementing all that you have learned: creating and choosing algorithms, running models, and the trial and error of training models to an acceptable level of accuracy. This will not happen overnight, but it will be a combination of hard work and perseverance using all the knowledge gained from the above and much more. It will be an advancement that will include many possibilities and insights. By preparing yourself at the start of the journey by answering the five questions within this white paper, you will find yourself well on the way to a successful and smoother experience with machine learning.

Leveraging the ability to transform data into actionable insights and proactive decisions is the key to competitive advantage for any asset and data-intensive organization. The ability to learn and evolve as new data is introduced – without explicitly programming to do so – would be the pinnacle of excellence in operations and reliability. That's what machine learning offers if done correctly: a capability that accelerates and extends data-driven insights and knowledge acquisition.

About Bentley's Machine Learning

Bentley's approach to getting the most out of machine learning is to work with users every step of the way to maximize its full potential. This is done by having a flexible approach to fitting the machine learning solution around the challenges and expected business outcomes and not forcing the solution around the problem and hope it works. By working together, AssetWise machine learning also leverages the neural network capabilities from the Bentley Institute to provide a wide range of solutions. Within AssetWise Asset Reliability and Operational Analytics, Bentley provides an exhaustive range of solutions that will extend our users' abilities to enable them to continue to achieve improved asset and operational reliability and optimal performance by providing smarter and faster ways of delivering situational intelligence and data-driven decision making across the whole operation.

References

1. Moore, S. (n.d.). How to do Machine Learning Without Hiring Data Scientists. Retrieved from https://www.gartner.com/smarterwithgartner/how-to-do-machine-learning-without-hiring-data-scientists/